

動詞概念と名詞概念の共起関係に基づく 事象における名詞概念の選択

—統合物語生成システムにおけるストーリー生成のための機構—

Noun Concepts Selection in Events Based on the Co-occurrence Relation in Verb Concepts and Noun Concepts: Mechanisms for the Story Generation in an Integrated Narrative Generation System

小野 淳平†
Jumpei Ono

小方 孝‡
Takashi Ogata

1. はじめに

統合物語生成システム (Integrated Narrative Generation System; INGS) [1,2]は, 事象概念の生成において, 動詞概念の格構造に伴う人物・物・場所を生成する際に, 名詞概念辞書より名詞概念の選択を行う. 本稿ではその選択処理に対して, 動詞概念における名詞概念の共起関係および名詞概念どうしの共起関係を利用する方法を導入する. なお, ここではあくまでも物語の概念レベルにおける構造の生成に関して扱っており, 表層レベルの表現に関しては扱っていない. 我々の文生成に関する研究は, 別稿[3, 4]に記す.

2. 背景

INGS (図 1) は物語の内容の構造を生成するストーリー機構, および語りの構造を生成する物語言説機構, 文字や音, 映像を媒体とした表現を生成する物語表現機構を持つ. これらの機構は我々の研究の成果を統合した機構である. また, 物語の内容や言語的表現を生成するための知識を格納する辞書, 静的な情報と動的な情報の相互関係を示す知識を格納する状態-事象変換知識ベース, 事象同士の因果関係を示す知識を格納する物語コンテンツ知識ベースからの知識を利用する. 辞書は概念辞書と言語表記辞書に分かれ, 概念と言語表記は1対1の関係を持つ.

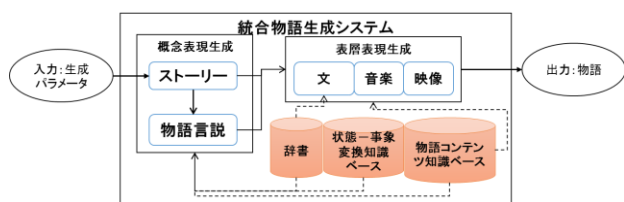


図 1 統合物語生成システムの構成図

概念辞書は, 名詞概念辞書と動詞概念辞書を持つ[5]. それぞれ意味的分類を示す中間概念と, 物語の具体的な要素を示す末端概念で構成される. 動詞概念は, 言語表現における文の基本形を示す“文型パターン”, 動詞概念が取る格を示す“格構造”, 各々の格が取り得る名詞概念の条件を示す“制約条件”を持つ. “制約条件”は名詞概念辞書に格納された中間概念により記述される (図 2). 概念辞書には, 他に形容詞・形容動詞・副詞に関する辞書も持つがここでは省略する.

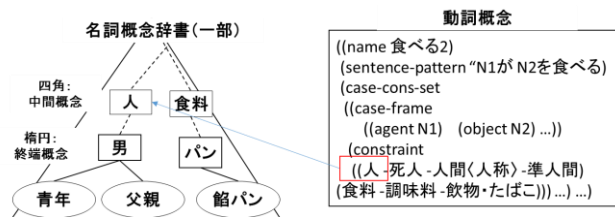


図 2 名詞概念辞書と動詞概念の制約条件の関係

INGS で扱う物語構造における個々の事象は上で述べた概念辞書に格納された, 一つの動詞概念とその格構造に伴う一つ以上の名詞概念 (人・物・場所等) として記述される. これを事象概念と呼ぶ. 事象概念生成では, 物語の内容を構成するための形式的手法であるストーリー技法で指定された動詞概念の格構造を元に, その動詞概念の制約条件に基づき候補の中から名詞概念がランダムに一つ選択される. その後, 選ばれた名詞概念を元に, 人物などの性質や特徴を記録する属性フレームを概念辞書から参照し, 物語を構成する具体的な人・物・場所を生成する. しかし, ある中間概念における末端概念の内容は均一ではないため, 選択される名詞概念において, その難しさや使用されている年代などの特徴にばらつきが生じていた. その問題に対して別稿[6]では, 「青空文庫」より得た名詞概念の出現頻度情報を利用した概念選択を試みた. 本稿では, 動詞概念と名詞概念の共起関係及び名詞概念どうしの共起関係を用いて, 制約範囲における名詞概念を絞り込む方法を提案・実装する.

3. 共起情報に基づく名詞概念の選択

この節では二種類の共起情報の算出方法について述べる. 共起情報の算出には「KH Coder」[7]を用いる. 「KH Coder」とはテキストデータを統計的に分析するソフトウェアである. また各方法には共通の素材として「青空文庫」収録の1872年から1963年までの新字新仮名の全作品4980作品 (2014年9月時点. 小説が中心. 他, 評論など含む. 本文部分のみを利用) から共起情報を算出した.

3.1 動詞概念における名詞概念の共起情報

ある動詞概念が持つそれぞれの格に対して, その動詞概念と共起関係を持つ名詞概念から一つを選択する方法について述べる. 例えば図3は各々の格で動詞概念との共起関係が最も強い名詞概念を選択している例である.

†岩手県立大学大学院 Graduate School of Iwate Prefectural University

‡岩手県立大学 Iwate Prefectural University

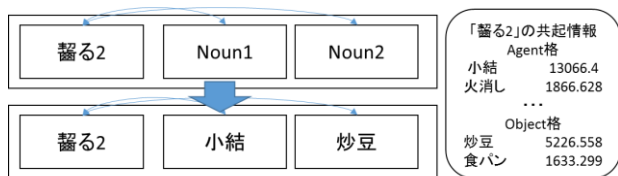


図 3 動詞概念と名詞概念の共起関係を利用した選択

3.1.1 動詞概念-名詞概念間の共起情報の算出方法

動詞概念における名詞概念の共起情報は、動詞概念辞書に格納された全 11951 個の動詞概念について算出する。

その算出手続きは、(A)動詞概念に対して共起関係を持つ名詞概念の収集、(B)獲得した共起情報と制約条件の比較による格ごとの共起情報の取捨選択、という二段階で構成される。

処理(A)では、全ての動詞概念毎に以下の処理を行う。

- (1)「KH Coder」の『関連語検索機能』を呼び出し、動詞概念の記述名から末尾の数字を消した語を検索する(例:『食べる 1』や『食べる 2』を『食べる』で検索)。
- (2)取得した名詞概念と共起情報の値の組をテキストファイルに記録する。このとき『●●する』形式の動詞概念(例:『料理する』など)は、構文解析器の都合上、『●● and する』と検索する。ここでは実質 4885 個の動詞概念について共起情報が算出される。

処理(B)では、上記の処理で共起情報を得た全ての動詞概念について、末尾に 1 から順に数字をつけ、該当する動詞概念が無くなるまで以下の処理を繰り返す。(1) 全ての格ごとに、共起関係を持つ名詞概念と制約条件を比較する、(2)制約条件に何れかの名詞概念が含まれる場合、それを抽出する、(3)全ての名詞概念を比較し終えたならば、格ごとに、抽出した名詞概念を要素としたリストを作る。

以上の二段階の処理の結果、各々の動詞概念について格ごとに制約条件の範囲で、共起関係を持つ名詞概念がその関係の強さと共に算出される。

3.1.2 動詞概念-名詞概念間の共起情報の算出の結果

上記の処理で得た結果を表 1 に示す。これは対象とした各々の動詞概念における格構造の全てを通して、一つ以上の名詞概念と共起関係を持つ格の個数を示している。

表 1 格単位で見た場合の算出結果

種類	該当数
共起関係を持つ名詞概念が有る格	31125
共起関係を持つ名詞概念が無い格	812

次に格の種類ごとに関する算出結果を表 2 に示す。要素の数とは、動詞概念と共起関係を持つ名詞概念のうち、幾つもの名詞概念が格の制約条件を満たしているかを示す。また表の最下段は、格の種類ごとに平均要素数の一割を満たしている格がどの程度あったかを示している。どの格においても、格の数の半数程度がそのラインを超えている。加えて、要素の数が 10 以上の格の数は全体の 7 割程度を占めており、十分に名詞概念の選択が行える程度の共起情報を得る事ができたと考える。

表 2 共起情報の算出結果のまとめ

	Agent	counter-agent	object	location	instrument	from	to
格の数	8734	3459	5259	9780	1709	963	1829
要素の数の平均	469.76	1064.56	1020.42	374.42	763.37	884.9	784.41
要素の数の標準偏差	595.89	1071.49	1597.18	633.58	2020.37	1634.13	1912.79
要素の数の最大	11347	21890	28312	28386	32589	19421	20313
要素の数の最少	0	0	0	0	0	0	0
要素平均の割を満たす格	7484	2521	2081	5769	844	525	809

共起関係を持つ名詞概念を持たない格の原因は以下の三通りである。

- (a)動詞概念が対象テキスト内に存在しない
- (b)制約条件が狭い
- (c)得た共起関係を持つ名詞概念が少ない

一つ目の原因は、451 個の動詞概念の格が該当する。これらは複合動詞(『巻き戻す』など)や、『～せる』形式の同士(『休ませる』など)である。根本的に存在しないデータとなるので、対象テキストを増やす必要がある。二つ目は制約条件が狭いため共起した名詞概念が一致しない事例である。極端な場合、終端概念が直接指定されている。例えば、動詞概念『上がる 19』の一つ目の格の制約条件は終端概念『遺体』となっている。この場合、選択肢が存在しないため、共起情報が介在する余地がない。終端概念が指定されていない場合でも、制約条件が示す中間概念が概念辞書の末端構造に近い。つまり、ある程度の具体性があり内部のばらつきが小さいカテゴリとなる。これらは選択される名詞概念の性質のばらつきを抑える目的から見た場合、共起情報を利用する必要性が少ない。三つ目は、共起関係を持つ名詞概念の数が他の動詞概念と比較しての一割程度の数で、制約条件を満たすことができなかった例が含まれる。素材の量を増やすだけでなく、名詞概念辞書へ新規に共起関係を持つ名詞概念を登録する必要もある。

3.2 名詞概念どうしの共起情報

この節では、ある名詞概念に対して共起関係を持つ名詞概念の中から一つを選択する方法について述べる。例えば図 4 は各々の格で共起関係が最も強い名詞概念を選択している。

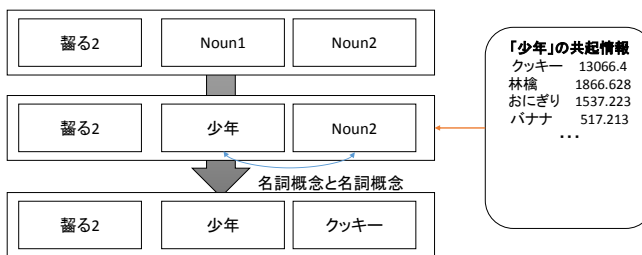


図 4 名詞概念どうしの共起情報に基づく選択

3.2.1 名詞概念どうしの共起情報の算出方法

名詞概念どうしの共起情報の算出では、総計 115765 の一般名詞の終端概念の内、76029 個が対象となる。省いた名詞概念は、『～学校』のように、『～』の位置に他の概念が接続されることで一つの概念となる「接辞に当たる概念」および異なる中間概念間で重複している終端概念である。基本的な手続きは動詞概念における名詞概念の共起情報の算出と同様に、「KH Coder」の『関連語検索機能』を呼びだし名詞概念を検索する。得た結果を、その共起関係の強さの情報と共に記録する。

3.2.2 名詞概念どうしの共起情報の算出の結果

表3に共起情報の算出結果を示す。平均要素数とは、一つの名詞概念が平均して、どの程度の数の名詞概念と共起関係を持つかを示しており、最大要素数はその数の最大値を、最少要素数は最小値を示す。

表3 共起情報の算出結果のまとめ

対象名詞概念総数	平均要素数	最大要素数	最少要素数
76029	37.6	70	0

要素数が0の名詞概念について、その原因は以下の二通りである。

- (a) 対象とした名詞概念が素材の中に存在しない
- (b) 共起した名詞概念が名詞概念辞書に未登録である

前者については、動詞概念を扱う場合と同様に、素材に存在しないデータである。素材の量を増やす必要がある。後者については、共起情報の抽出段階で得た名詞概念を、名詞概念辞書に追加する作業を必要とする。その場合、名詞概念の意味的な分類を行う必要があり、大掛かりな作業となり得る。

4. アンケート調査

この節では、3節で示した共起情報の算出結果に基づく物語の生成方法を示し、実際に生成した物語の評価とその考察を行うことで、共起情報に基づく生成では生成され物語に対して、どのような効果があるのか確認する。

4.1 共起情報を利用した物語の生成方法

2節で示した動詞概念の共起情報の結果を利用した名詞概念の選択手法について述べる。動詞概念の決定までは現状の INGS と同じ方法であり、名詞概念辞書から名詞概念を一つ選択する手続きを拡張する。

入力には、名詞概念の選択基準と動詞概念の格構造であり、出力は動詞概念の格構造の全ての格に対応する名詞概念である。共起情報は処理の途中で参照する。名詞概念の選択基準は、選択される名詞概念がどの程度、動詞概念ないしは直前の格に挿入された名詞概念と共起関係を持つかを指定する。ここでは、最も共起関係が強い概念を選択する『最大』と、最も共起関係が弱い概念を選択する『最少』のどちらかを指定する。

動詞概念と名詞概念の共起情報を参照する場合は、動詞概念が持つ格の数だけ、格ごとに存在する共起情報のリストを参照し、名詞概念を一つ選択する。選択候補が

複数ある場合は、その中からランダムに選択する。名詞概念どうしの共起情報を参照する場合は、その動詞概念における最初の格はランダムに決定する。続く格に関しては、その直前の格が持つ名詞概念の共起情報のリストを参照し、制約条件の範囲で名詞概念を一つ選択する。選択候補が複数ある場合は、上の方法と同様である。

共起情報に基づく名詞概念の選択による物語を図5に例示する。この例では動詞概念と名詞概念の共起情報を用いて、共起関係が最も強い名詞概念の選択を行った。図の上段が、生成された物語であり、図の下段は言語表記辞書の代表表記を用いて、最も単純な形で生成した文表現である（概念構造の下線が引かれた部分が、文表現の下線が引かれている部分と対応）。物語の全体の構造は、我々が開発したストーリーコンテンツグラマー[8]を利用して生成した。なお1節で述べた通り、本稿ではあくまで概念構造の生成に着目しており、文表現は最低限の見易さを求めるために示している。

概念構造	(\$ロシア魔法昔話 (\$問題 (\$予備部分 (\$01_留守 (event 出かける 1 (type action) (ID 1) (time (time1 time2)) (agent age%三后#1) (location loc%湖畔#1) (object (event 摘む 1...)) ...)) …<以下省略>…)
	三后が「三后が練綿を摘む」ために出かける。単細胞が良剤を良剤に混入する。単細胞が良剤を三后に与える。三后が眠る。単細胞が鉄工に「鉄工が三后を百味筆筒に詰める」ことを命令する。鉄工が三后を百味筆筒に詰める。…<以下、省略>…
文表現	

図5 共起情報に基づく概念選択による生成例

4.2 調査の手続き

二種類の共起情報に基づく物語について、共通の調査手続きを利用し、それぞれ独立して評価を行った。

評価では、生成した物語から名詞概念だけを抜き出した名詞概念の一覧と、生成した物語を単文単位でばらにした文のリストを評価対象として調査する。

評価対象は二種類の共起情報ごとに、二段階の過程で作成している。まず、『最大』の選択、『最少』の選択、および制約条件の範囲内で概念のランダム選択（『乱数』と呼ぶ）、以上の三種類の選択方法で各100個ずつ、合計300個の物語を生成した。次に、それぞれの方法ごとに無作為に2個ずつ物語を選択し、計6個の物語から評価対象を作成する。名詞概念の一覧については、それぞれの物語から名詞概念部分を抽出した。文のリストは、物語を文生成機構で、文表現に変換し、それぞれ単文単位で分割して作成した。何れの評価対象も三種類の選択方法のそれぞれについてランダムに混ぜ合わせ、名詞概念の一覧および文のリストを一つずつ作成している。

評価の手順は次の通りである一被験者（評価者）は、紙に印刷された名詞概念の一覧および文のリストを読み進め、名詞概念が現れるたびに、段階で1（易しい）から5（難しい）までの何れかの評価を行う。このとき、名詞概念「男」を評価値2とし、それよりも簡単な名詞概念に対して1を、それよりも難しい名詞概念に対しては、その難しさに応じて3以上の値をつけさせた。その際、名詞概念の一覧については5分、文のリストについては7分の制限時間内で作業を終える。一度付けた評価は訂正しな

い・前と同じ名詞概念が現れた場合前回の評価を確認しない(評価が違って構わない)等の条件を付す。評価者は、大学生8名であり、名詞概念の一覧については3セット分、文のリストについては2セット分の評価を行った。

4.3 調査結果と考察

以下の表4にアンケート調査の結果を示す。上段が動詞概念と名詞概念の共起情報側の結果であり、下段が名詞概念どうしの共起情報側の結果である。

表4 共起情報に基づく物語に関する調査結果

組み合わせ	種類	最大	最少	乱数
動詞概念—名詞概念	名詞概念のみ-1回目	2.54	2.05	2.67
	名詞概念のみ-2回目	2.49	1.86	2.28
	名詞概念のみ-3回目	3.27	2.39	2.81
	文-1回目	2.32	1.78	2.17
	文-2回目	2.79	2.04	2.17
	平均値	2.68	2.02	2.42
名詞概念—名詞概念	名詞概念のみ-1回目	2.48	1.78	2.31
	名詞概念のみ-2回目	2.06	1.67	2.06
	名詞概念のみ-3回目	3.13	2.44	2.84
	文-1回目	2.34	1.75	2.17
	文-2回目	2.18	1.74	2.08
	平均値	2.44	1.88	2.29

アンケート調査では、二種類の共起情報のどちらにおいても、共起関係が強い概念に基づく生成を行った場合、難しい概念が選択されたという結果を得た。より詳しく結果を見る。まずアンケート結果のみで判断できる、名詞概念どうしの共起情報を利用した生成において『最大』の選択における評価値が上位および下位3つまでの評価結果を表5と表6に示す。評価項目は実際に評価した名詞概念を示し、平均は被験者ごとの評価結果の平均値を示す。同時に選択された名詞概念とは、一つの事象概念の中で、評価項目の名詞概念と共起関係を持つ名詞概念として選択された名詞概念の一例を示している。これらの名詞概念は、半数以上の被験者が3以上の値を付けている。また『最少』の選択の評価結果を見た場合、評価値1.63の『男』は評価平均値3.25の『内親王』が選択され、評価平均値3.88の『臥所』からは、評価平均値2.00の『勇士』が選択されている。

表5 難しいと判断された名詞概念(『最大』)

評価項目	評価平均	同時に選択された名詞概念
歿	4.625	大殿(評価平均2.88)
割栗石	4.375	労務(評価平均2.62)
先夫	4	壮士(評価平均3.37)

表6 易しいと判断された名詞概念(『最大』)

評価項目	評価平均	同時に選択された名詞概念
船	1.3	父(評価平均1.38)
血	1.3	鉄塔(評価平均2.00)
馬	1.5	勇士(評価平均2.00)

次に動詞概念と名詞概念の共起情報の場合を、表7と表8に示す。表の項目のうち、評価対象および評価平均は表

5と同一である。共起関係を持つ動詞概念は、名詞概念の選択の基準となった動詞概念を示しており、括弧内の『最少』および『最大』は選択方法をそのまま表している。『買う』『落とす』は比較的簡単な動詞概念として、『出現する』『出国する』は比較的難しい動詞概念として示す。この難易度の判断は「リーディングチュウ太」[9]と呼ばれる日本語読解学習支援システムを利用した。

表7 難しいと判断された名詞概念

評価対象	評価平均	共起関係を持つ動詞概念
龐大	4.62	買う(『最少』)
補陀落	4.63	出現する(『最大』)

表8 易しいと判断された名詞概念

評価対象	評価平均	共起関係を持つ動詞概念
指輪	1.5	落とす(『最大』)
山	1.25	出国する(『最少』)

以上の事から難しい概念と共起関係が強い概念は難しい概念となり得る可能性が高いと仮定できる。また難しい概念と共起関係が弱い概念は易しい概念となる。また易しい概念を基準に考えた場合、その逆が考えられる。この結果から概念選択における概念の難しさの制御に対して、効果的な共起情報の利用が考えられる。

5. おわりに

本稿では、動詞概念と名詞概念および名詞概念どうしの共起関係を抽出し、それを利用した名詞概念の選択方法を提案した。また実際に生成した物語に含まれる名詞概念に関して、その易しさ/難しさの評価実験を行い、難しい概念と共起関係が強い概念は難しい概念である(またはその逆)という仮説を得た。今後はこの仮説を検証するための実験を行い、共起関係を利用した概念の難しさを制御する方法に関して提案を行っていく。

参考文献

- [1]小方孝, 金井明人, “物語論の情報学序説—物語生成の思想と技術を巡って—”, 学文社 (2010).
- [2]Akimoto, T. and Ogata, T., “An Information Design of Narratology: The Use of Three Literary Theories in a Narrative Generation System”, The International Journal of Visual Design, 7(3), 31-61, (2014).
- [3]小野淳平, 小方孝 “統合物語生成システムにおける概念選択/語彙表記選択及びその制御”, 第29回人工知能学会全国大会論文集, 3G4-OS-05a-3 (2015).
- [4]小方孝, 小野淳平 “統合物語生成システムにおける言語表記辞書とその利用”, 信学技報, 115(69), 25-30 (2015).
- [5]Ogata, T., “Building Conceptual Dictionaries for an Integrated Narrative Generation System”, Journal of Robotics, Networking and Artificial Life, 1(4), 270-284 (2015).
- [6]小野淳平, 小方孝, “計量データに基づく名詞概念の選択—「統合物語生成システム」における一機構として—”, 信学技報, 114(366), 49-54 (2014).
- [7]樋口耕一, “テキスト型データの計量的分析—2つのアプローチの峻別と統合—”, 理論と方法, 19(1), 101-115 (2004).
- [8]Imabuchi, S. and Ogata, T., “A Story Generation System Based on Propp Theory: As a Mechanism in an Integrated Narrative Generation System”, Lecture Note of Artificial Intelligence, Springer, 7614, 312-321 (2012).
- [9]Kawamura, Y., Kitamura, T. and Hobara, R., “リーディングチュウ太”, <http://language.tiu.ac.jp/> (最終アクセス2015年) .